NeoBiota 44: 1–25 (2019)
doi: 10.3897/neobiota.44.31650
http://neobiota.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
NeoBiota
Advancing research on alien species and biological invasions

Consistency of impact assessment protocols for non-native species

Pablo González-Moreno[1], Lorenzo Lazzaro[2], Montserrat Vilà[3], Cristina Preda[4,5], Tim Adriaens[6], Sven Bacher[4], Giuseppe Brundu[7], Gordon H. Copp[8,9], Franz Essl[10], Emili García-Berthou[11], Stelios Katsanevakis[12], Toril Loennechen Moen[13], Frances E. Lucy[14], Wolfgang Nentwig[15], Helen E. Roy[16], Greta Srėbalienė[17], Venche Talgø[18], Sonia Vanderhoeven[19], Ana Andjelković[20,21], Kęstutis Arbačiauskas[22], Marie-Anne Auger-Rozenberg[23], Mi-Jung Bae[11,24], Michel Bariche[25], Pieter Boets[26], Mário Boieiro[27], Paulo Alexandre Borges[27], João Canning-Clode[28,29,30], Frederico Cardigos[31], Niki Chartosia[32], Elizabeth Joanne Cottier-Cook[33], Fabio Crocetta[34], Bram D'hondt[35], Bruno Foggi[2], Swen Follak[36], Belinda Gallardo[37], Øivind Gammelmo[38], Sylvaine Giakoumi[39], Claudia Giuliani[40], Guillaume Fried[41], Lucija Šerić Jelaska[42], Jonathan M. Jeschke[43,44,45], Miquel Jover[46], Alejandro Juárez-Escario[47,48], Stefanos Kalogirou[49], Aleksandra Kočić[50], Eleni Kytinou[51], Ciaran Laverty[52], Vanessa Lozano[7], Alberto Maceda-Veiga[3], Elizabete Marchante[53], Hélia Marchante[53,54], Angeliki F. Martinou[55], Sandro Meyer[56], Dan Michin[57,58], Ana Montero-Castaño[3], Maria Cristina Morais[53,59], Carmen Morales-Rodriguez[60], Nadia Muhthassim[15], Zoltán Á. Nagy[61], Nikica Ogris[62], Huseyin Onen[63], Jan Pergl[64], Riikka Puntila[65], Wolfgang Rabitsch[66], Triya Tessa Ramburn[67], Carla Rego[27], Fabian Reichenbach[15], Carmen Romeralo[68,69], Wolf-Christian Saul[43,44,45], Gritta Schrader[70], Rory Sheehan[14], Predrag Simonović[71], Marius Skolka[5], António Onofre Soares[72], Leif Sundheim[18], Ali Serhan Tarkan[73], Rumen Tomov[74], Elena Tricarico[2], Konstantinos Tsiamis[75], Ahmet Uludağ[76], Johan van Valkenburg[77], Hugo Verreycken[78], Anna Maria Vettraino[79], Lluís Vilar[46], Øystein Wiig[80], Johanna Witzell[69], Andrea Zanetta[4,81], Marc Kenis[82]

1 *CABI, Egham, UK* **2** *Department of Biology, University of Florence, Florence, Italy* **3** *Estación Biológica de Doñana (EBD-CSIC), Sevilla, Spain* **4** *University of Fribourg, Department of Biology, Fribourg, Switzerland* **5** *Ovidius University of Constanta, Department of Natural Sciences, Constanta, Romania* **6** *Research Institute for Nature and Forest (INBO), Brussels, Belgium* **7** *Department of Agriculture, University of Sassari, Sassari, Italy* **8** *Salmon & Freshwater Team, CEFAS, Lowestoft, UK* **9** *Centre for Conservation Ecology and Environmental Science, Bournemouth University, Poole, UK* **10** *Division of Conservation, Vegetation and Landscape Ecology, University Vienna, Vienna, Austria* **11** *GRECO, Institute of Aquatic Ecology, University of Girona, Girona, Spain* **12** *University of the Aegean, Department of Marine Sciences, Mytilene, Greece* **13** *Norwegian Biodiversity Information Centre. Trondheim. Norway* **14** *CERIS, Institute of Technology, Sligo, Ireland* **15** *In-*

Copyright *Pablo González-Moreno et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

stitute of Ecology and Evolution, University of Bern, Bern, Switzerland **16** *Centre for Ecology & Hydrology, Crowmarsh Gifford, UK* **17** *Marine Science and Technology Centre, Klaipėda University, Klaipėda, Lithuania* **18** *Norwegian Institute of Bioeconomy Research (NIBIO), Ås, Norway* **19** *Belgian Biodiversity Platform, Walloon Research Department for Nature and Agricultural Areas (DEMNA), Service Public de Wallonie, Gembloux, Belgium* **20** *Institute for Plant Protection and Environment, Belgrade, Serbia* **21** *Department of Biology and Ecology, Faculty of Sciences, University of Novi Sad, Novi Sad, Serbia* **22** *Nature Research Centre, Akademijos Street 2, LT-08412 Vilnius, Lithuania* **23** *INRA, UR633, Zoologie Forestière (URZF), Orléans, France* **24** *Freshwater Biodiversity Research Division, Nakdonggang National Institute of Biological Resources, Gyeongsangbuk-do 37242, Republic of Korea* **25** *Department of Biology, American University of Beirut, Beirut, Lebanon* **26** *Provincial Centre of Environmental Research (PCM), Ghent, Belgium* **27** *cE3c – Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group and Universidade. dos Açores – Depto de Ciências e Engenharia do Ambiente, Azores, Portugal* **28** *MARE – Marine and Environmental Sciences Centre, Madeira Island, Portugal* **29** *Centre of IMAR of the University of the Azores, Department of Oceanography and Fisheries. Horta, Azores, Portugal* **30** *Smithsonian Environmental Research Center, Edgewater, MD 21037, USA.* **31** *OKEANOS - Research Center – Universidade dos Açores, Azores, Portugal* **32** *Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus* **33** *Scottish Association for Marine Science, Scottish Marine Institute, Oban, UK* **34** *Department of Integrative Marine Ecology, Stazione Zoologica Anton. Dohrn, Villa Comunale, I-80121 Napoli, Italy* **35** *Biology Department, Ghent University, Ghent, Belgium* **36** *Austrian Agency for Health and Food Safety, Institute for Sustainable Plant Production, Vienna, Austria* **37** *Applied and Restoration Ecology Group (IPE-CSIC), Zaragoza, Spain* **38** *BioFokus, Oslo, Norway* **39** *Université Côte d'Azur, CNRS, UMR 7035 ECOSEAS, Nice, France* **40** *Department of Pharmaceutical Sciences (DISFARM), University of Milane, Milane, Italy* **41** *Plant Health Laboratory, Anses, Montferrier-sur-Lez, France* **42** *Department of Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia* **43** *Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany* **44** *Freie Universität Berlin, Department of Biology, Chemistry, Pharmacy, Institute of Biology, Berlin, Germany* **45** *Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), Berlin, Germany* **46** *Unitat de Botànica, Facultat de Ciències, Campus de Montilivi, University of Girona, Girona, Spain* **47** *Department of Horticulture, Fruit Growing, Botany and Gardening, Agrotecnio, ETSEA, University of Lleida, Spain* **48** *Department of Forest and Crop Science, Agrotecnio, ETSEA, University of Lleida, Spain* **49** *Hellenic Centre for Marine Research, Hydrobiological Station of Rhodes, Rhodes, Greece* **50** *Department of Biology, Josip Juraj Strossmayer University of Osijek, Osijek, Croatia* **51** *Department of Marine Sciences, University of the Aegean, Lesvos Island, Greece* **52** *School of Biological Sciences, Medical and Biological Centre, Queen's University Belfast, UK* **53** *Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Coimbra, Portugal* **54** *Instituto Politécnico de Coimbra, Escola Superior Agrária de Coimbra, Coimbra, Portugal* **55** *Joint Services Health Unit, RAF Akrotiri, British Forces Cyprus, Cyprus* **56** *Department of Environmental Sciences, University of Basel, Basel, Switzerland* **57** *Marine Science and Technology Centre, Klaipeda University, Klaipeda, Lithuania* **58** *Marine Organism Investigations, Ballina, Killaloe, Ireland* **59** *Centre for the Research and Technology of Agro-Environmental and Biological Sciences, Department of Biology and Environment, University of Tras-os-Montes and Alto Douro, Vila Real, Portugal* **60** *Pathology of Woody Plants. Technische Universität München, TUM, Freising, Germany* **61** *Phytophthora Research Centre, Department of Forest Protection and Wildlife Management, Faculty of Forestry and Wood Technology, Mendel University in Brno; Brno, Czech Republic* **62** *Slovenian Forestry Institute, Ljubljana, Slovenia* **63** *Department of Plant Protection, Faculty of Agriculture, Gaziosmanpasa University, Tokat, Turkey* **64** *Department of Invasion Ecology, Institute of Botany, The Czech Academy of Sciences, Průhonice, Czech Republic* **65** *Marine Research Centre, Finnish Environment Institute, Helsinki, Finland* **66** *Environment Agency Austria, Vienna, Austria* **67** *Simon Fraser University, Burnaby, Canada* **68** *Sustainable Forest Management Research Institute, University of Valladolid-INIA, Palencia, Spain* **69** *Swedish University of Agricultural Sciences, Faculty of Forest Sciences, Southern Swedish Forest Research Centre, Alnarp, Sweden* **70** *Julius Kuehn Institute (JKI), Braunschweig,*

Germany **71** *Faculty of Biology & Institute for Biological Research "Siniša Stanković", University of Belgrade, Belgrade, Serbia* **72** *cE3c – Centre for Ecology, Evolution and Environmental Changes/Azorean Biodiversity Group and University of the Azores – Faculty of Sciences and Technology, Açores, Portugal* **73** *Faculty of Fisheries, Muğla Sıtkı Koçman University, Muğla, Turkey* **74** *University of Forestry, Department of Plant Protection, Sofia, Bulgaria* **75** *European Commission, Joint Research Centre (JRC), Ispra, Italy* **76** *Faculty of Agriculture, Çanakkale Onsekiz Mart University, Çanakkale, Turkey* **77** *National Plant Protection Organization (NPPO), Wageningen, The Netherlands* **78** *Research Institute For Nature and Forest (INBO), Linkebeek, Belgium* **79** *Department for Innovation in Biological, Agro-food and Forest systems, University of Tuscia, Viterbo, Italy* **80** *Natural History Museum, University of Oslo, Oslo, Norway* **81** *Swiss Federal Research Institute WSL, Biodiversity and Conservation Biology, Birmensdorf, Switzerland* **82** *CABI, Delémont, Switzerland*

Corresponding author: *Pablo González-Moreno* (p.gonzalez-moreno@cabi.org)

Academic editor: *P. Hulme* | Received 14 November 2018 | Accepted 26 February 2019 | Published 1 April 2019

Citation: González-Moreno P, Lazzaro L, Vilà M, Preda C, Adriaens T, Bacher S, Brundu G, Copp GH, Essl F, García-Berthou E, Katsanevakis S, Moen TL, Lucy FE, Nentwig W, Roy HE, Srėbalienė G, Talgø V, Vanderhoeven S, Andjelković A, Arbačiauskas K, Auger-Rozenberg M-A, Bae M-J, Bariche M, Boets P, Boieiro M, Borges PA, Canning-Clode J, Cardigos F, Chartosia N, Cottier-Cook EJ, Crocetta F, D'hondt B, Foggi B, Follak S, Gallardo B, Gammelmo Ø, Giakoumi S, Giuliani C, Fried G, Jelaska LS, Jeschke JM, Jover M, Juárez-Escario A, Kalogirou S, Kočić A, Kytinou E, Laverty C, Lozano V, Maceda-Veiga A, Marchante E, Marchante H, Martinou AF, Meyer S, Michin D, Montero-Castaño A, Morais MC, Morales-Rodriguez C, Muhthassim N, Nagy ZA, Ogris N, Onen H, Pergl J, Puntila R, Rabitsch W, Ramburn TT, Rego C, Reichenbach F, Romeralo C, Saul W-C, Schrader G, Sheehan R, Simonović P, Skolka M, Soares AO, Sundheim L, Tarkan AS, Tomov R, Tricarico E, Tsiamis K, Uludağ A, van Valkenburg J, Verreycken H, Vettraino AM, Vilar L, Wiig Ø, Witzell J, Zanetta A, Kenis M (2019) Consistency of impact assessment protocols for non-native species. NeoBiota 44: 1–25. https://doi.org/10.3897/neobiota.44.31650

Abstract

Standardized tools are needed to identify and prioritize the most harmful non-native species (NNS). A plethora of assessment protocols have been developed to evaluate the current and potential impacts of non-native species, but consistency among them has received limited attention. To estimate the consistency across impact assessment protocols, 89 specialists in biological invasions used 11 protocols to screen 57 NNS (2614 assessments). We tested if the consistency in the impact scoring across assessors, quantified as the coefficient of variation (CV), was dependent on the characteristics of the protocol, the taxonomic group and the expertise of the assessor. Mean CV across assessors was 40%, with a maximum of 223%. CV was lower for protocols with a low number of score levels, which demanded high levels of expertise, and when the assessors had greater expertise on the assessed species. The similarity among protocols with respect to the final scores was higher when the protocols considered the same impact types. We conclude that all protocols led to considerable inconsistency among assessors. In order to improve consistency, we highlight the importance of selecting assessors with high expertise, providing clear guidelines and adequate training but also deriving final decisions collaboratively by consensus.

Keywords

Environmental impact, expert judgement, invasive alien species policy, management prioritization, risk assessment, socio-economic impact

Introduction

Coupled with the increasing evidence of adverse impacts exerted by some non-native species (NNS) on native species and ecosystems (Katsanevakis et al. 2014, Vilà et al. 2011, Vilà and Hulme 2017), there is an increasing demand for robust and user-friendly impact assessment protocols to be used by professionals with different levels of expertise and knowledge. Such protocols are needed to predict impacts of new or likely invaders as well as to assess the actual impact of established species. Scientists, environmental managers, conservationists, and policy makers are developing and implementing approaches to prevent further NNS introductions and their subsequent establishment, spread and impact. Risk analysis associated with these four main phases of the invasion process is used to inform management decisions, such as whether to eradicate or control species that arrive despite prevention efforts (Leung et al. 2012). Assessment of the realized or potential impacts of NNS is particularly important for the prioritization of management actions (Essl et al. 2011). However, the large variety of metrics adopted to measure the impacts undermines direct comparison of impacts across species, groups of taxa, localities or regions (Vilà et al. 2010). To this end, protocols to integrate and synthesize the empirical evidence of NNS impacts are needed in order to ensure a rational use of resources (McGeoch et al. 2016), or for prioritizing species for subsequent risk assessment (Brunel et al. 2010, Copp et al. 2009).

Robust NNS impact protocols should ideally result in accurate and consistent impact scores for a species even if applied by different assessors, as long as they have the adequate expertise in the assessed species and context. However, despite the importance of consistency in impact protocols, we have little understanding of the patterns in consistency of impact scores across assessors and protocols, and more importantly, which factors contribute to high levels of consistency. The level of consistency in species scores across assessors may depend on the characteristics of the protocol (e.g. taxonomic and environmental scope, impact types included), but also on the available scientific evidence of impact, and the level of expertise of assessors. For instance, we may expect high consistency (i.e. low impact score variability) across assessors for well-studied species, or when all assessors have an in-depth understanding of the species under consideration.

Several international and national organizations and research groups have developed NNS protocols (Table 1). The common aspect of most of these protocols is that they allow a ranking of NNS according to the threat they pose to the risk assessment area. These have been applied for identifying and assessing potential NNS impacts at different spatial scales, e.g. continental (Nentwig et al. 2010) or national (D'hondt et al. 2015). However, these protocols differ in several aspects. For example, they vary according to their objective, with some considering only environmental impacts whereas others are broader and include socio-economic or ecosystem services impacts (Leung et al. 2012, McGeoch et al. 2016, Vanderhoeven et al. 2017). Some protocols were designed to be taxonomically generic (e.g. GB-NNRA), whereas others are specific for the screening of certain taxonomic groups such as fish or other aquatic organisms (e.g. FISK, MI-ISK, FI-ISK, Amph-ISK, EPPO-PRI; see Table 1), particular habitats (e.g.

BINPAS), or pathways (Panov et al. 2009). Moreover, the existing protocols vary considerably in complexity, such as the number of questions, the need for peer review, the use of additional software (e.g. spreadsheet or online form), the ways of assessing uncertainty (Vanderhoeven et al. 2017), and the scoring system used, which can be categorical, ordinal or continuous (Roy et al. 2018). The content and structural differences among protocols could lead to differences in the assessment results (Leung et al. 2012).

A few comparative analyses have addressed differences in the structure of impact assessment protocols (Essl et al. 2011, Heikkilä 2011, Vilà et al. 2019), and on their consistency in ranking species across regions (Matthews et al. 2017). However, studies have focused on a reduced number of protocols, and a short list of species (Křivánek and Pyšek 2006, Turbé et al. 2017). An in-depth comparison across taxa and across standardized protocols is missing for Europe (Essl et al. 2011), or elsewhere (Snyder et al. 2013). Such a comparison is urgently required to respond to the European legislation on invasive NNS (Regulation EU No. 1143/2014). The aim of the present study was to test for consistency in assessment scores across assessors through comparison of several NNS impact assessment protocols. To address this aim, 89 invasive NNS specialists used 11 protocols to assess the potential impact of 57 species not native to Europe and belonging to a very large array of taxonomic groups (plants, animals, pathogens) from terrestrial to freshwater and marine environments. The specific questions considered were: 1) How consistent are species scores across assessors? 2) To what extent does consistency depend on the protocol characteristics, i.e. impact categories considered (environmental and socio-economic), structural complexity of the protocol (number of questions and scoring system)? 3) How is consistency related to the characteristics of the NSS (taxonomic group, habitat type, and available scientific knowledge of the species); 4) What is the relation between consistency and assessor expertise? 5) Do different protocols provide similar final scores or species ranking? Based on the study results, we provide recommendations on how the robustness and applicability of protocols could be improved for assessing NNS impacts.

Material and methods

Selection of impact assessment protocols

Eleven commonly used scientifically based protocols developed or applied in Europe for the evaluation of NNS impacts were selected for comparison by consensus in the AlienChallenge COST Action workshop in April 2014 by 36 European experts in NNS risk assessments (Rhodes, Greece) (Table 1). We included all protocols developed and officially used at national or continent level in Europe (e.g. EPPO, Harmonia+ and GB-NNRA) and the main protocols used by European research community (e.g. GISS and FISK). Only the EFSA protocol was discarded from this selection due to the complexity of extracting and processing the data. Furthermore, during the selection we aimed to cover the major types and groups of protocols in order to guarantee

Table 1. Characteristics of impact assessment protocols used in the study. Each protocol is characterized in terms of the a) taxonomic group the protocol could be used for, b) the impact categories included (environmental alone or environmental and socio-economic), c) the final scoring scale (i.e. three levels, five levels, and more than 5 levels), d) whether the final score is based on the maximum score of impacts, e) whether the protocol included questions on species spread as part of a risk assessment (yes/no), f) the number of questions contributing to the final score, and g) the mean assessor expertise on species required to fill the questionnaire (1–5 scale based on 63 online anonymous questionnaire responses).

Protocol	Full name	Taxonomic groups	Impact categories	Final scoring scale	Final scoring based on maximum score	Spread questions included	Number of questions	Expertise on species required	Reference
BINPAS	Biological Invasion Impact/Biopollution Assessment	Aquatic animals	Environmental	5	yes	yes	5	3.50	(Narščius et al. 2012, Olenin et al. 2007, Zaiko et al. 2011)
EICAT	Environmental Impact Classification for Alien Taxa	All	Environmental	5	yes	no	9	3.37	(Blackburn et al. 2011, Hawkins et al. 2015)
EPPO-EIA	European Plant Protection Organisation-Environmental Impact Assessment for plants (EPPO-EIA-PL) and terrestrial invertebrates (EPPO-EIA-IN)	Terrestrial plants and invertebrates	Environmental	5	yes	no	8 (Plants); 9 (invert.)	3.16	(Kenis et al. 2012)
EPPO-PRI	EPPO-Prioritization scheme	Plants	Environmental and socio-economic	3	yes	yes	11	3.00	(Brunel et al. 2010)
FISK (and related)	Fish Invasiveness Screening Kit (FISK); Freshwater Invertebrate Invasiveness Screening Kit (FI-ISK); Marine Fish Invasiveness Screening Kit (MFISK); Marine Invertebrate Invasiveness Screening Kit (MI-ISK)	Aquatic animals	Environmental and socio-economic	3	no	yes	49	4.12	(Copp 2013, Copp et al. 2009, Panov et al. 2009, Tricarico et al. 2010)
GABLIS	German-Austrian Black List Information System	All	Environmental	3	yes	yes	12	3.22	(Essl et al. 2011)
GB-NNRA	Great Britain Non-native Species Risk Assessment	All	Environmental and socio-economic	5	no	yes	33	3.90	(Baker et al. 2008, Mumford et al. 2010)
GISS	Generic Impact Scoring System	All	Environmental and socio-economic	>5 (discrete with max 60)	no	no	12	3.46	(Nentwig et al. 2010, 2016)
Harmonia+	Belgian risk screening tools for potentially invasive plants and animals	All	Environmental and socio-economic	>5 (continuous	yes	yes	20	3.46	(D'hondt et al. 2015)
ISEIA	Belgian Invasive Species Environmental Impact Assessment	All (not marine for this study)	Environmental	3	no	yes	4	2.81	(Branquart 2009)
NGEIAAS	Norway Generic Ecological Impact Assessment of Alien Species	All	Environmental	5	yes	yes	11	4.34	(Gederaas et al. 2012, Sandvik et al. 2013)

enough variability in their characteristics. The selection does not consider risk analysis tools or updates that have become available after 2015, such as AS-ISK (Copp et al. 2016), which replaces FISK and the other -ISK toolkits and complies with the minimum standards NNS risk analysis under Regulation (EU) No 1143/2014 (Roy et al. 2018). Risk assessments are usually divided into four components that consider the potential for a non-native species to enter a region, establish, spread and cause impacts. The selection included impact assessment and risk assessment protocols for which we only compared the sections dealing with spread and impact as they are largely interrelated. Each protocol considers a different method to calculate the final score per species based on the responses (i.e. aggregation method): maximum impact, accumulated impact, categorization matrix or decision trees, an independent summary question, or the combination of any of the previous methods. Owing to the number of protocols used in the present study and their complexity, no attempt was made to standardize variations in score aggregation methods but rather, where possible, to account for this variability during the data analysis as covariates. Some protocols can be applied to any taxon while others are specific to particular groups or habitats (e.g. BINPAS and FISK are used only for aquatic animals, EPPO Prioritization for plants). As such, the number of protocols assessed per species varied depending on the taxonomic group (Table 1). Although all the -ISK toolkits (FISK, FI-ISK, Amph-ISK, MFISK, MI-ISK) were used for their respective taxonomic groups, in the data analyses all the versions were listed under 'FISK' because of their high similarity. For the same reason, the EPPO-EIAs for insects/pathogens and plants were listed together.

Each protocol was characterized according to several variables (Table 1): the categories of impact considered (environmental alone or environmental and socio-economic), inclusion of questions on species spread (yes/no), on scoring scale (i.e. three levels, five levels and more than five levels), whether the protocol included a maximum aggregation method (i.e. the largest value of a set of values) to calculate the final score (yes/no), the number of questions requiring input from the assessors and contributing to the final score, and the expertise on the species required to complete the protocol. The latter was based on 63 responses received from an online anonymous questionnaire distributed to all assessors, which included a question asking them to rate their agreement (from 1 = disagree to 5 = fully agree) with the statement: "This protocol requires a high level of expertise on the species". Assessors answered this question for each protocol after having completed all assessments. The response values were averaged per protocol to provide a single estimate of the level of expertise required for that NNS protocol (Table 1).

Selection of species

A total of 57 species from different taxonomic groups not native to terrestrial, freshwater, and marine environments in Europe were selected (Suppl. material 1: Table S1). Among them, only two species are native to a part of Europe (*Arion vulgaris* and *Dreissena polymorpha*). The list of species was elicited by consensus also at the Al-

ien Challenge COST Action workshop in April 2014 (Rhodes, Greece). During the workshops, the experts were grouped according to their taxonomic expertise under the coordination of a taxonomic leader, in order to select a list of species covering a wide range of European climatic regions and habitat types, biological characteristics and the degree and type of impact. While some NNS were widespread, very well studied and with known impacts, some had a localized geographical distribution (Suppl. material 1: Table S1). Each NNS was assigned to a specific taxonomic group and habitat type: terrestrial plants, freshwater plants, terrestrial vertebrates, terrestrial insects, other terrestrial invertebrates, freshwater invertebrates, freshwater fish, marine species, and pathogens. The scientific knowledge available for the NNS was quantified as the number of records in the Web of Science using the accepted scientific name as a query, and biology and ecology research area as filters (retrieved in August 2016). Additionally, the mean and coefficient of variation of the assessor expertise on each species (Suppl. material 1: Table S1) was derived through a self-valuation questionnaire on each assessed NNS using the following classification: 1 = low (the assessor has not worked with the species); 2 = medium (the assessor has not published on the species but has expertise on it through surveys or reports); and 3 = high (the assessor has published on the species).

Assessment of non-native species

There is a large variation in methods to implement the different protocols; some are available as downloadable freeware (-ISK toolkits, the 'NAPRA' version of the GB-NNRA), as online applications (e.g. Harmonia+, BINPAS), whereas some have to be constructed following the text guidelines (e.g. GISS, EICAT), and others can be obtained as spreadsheets (e.g. GB-NNRA) or databases (e.g. NGEIAAS). To harmonize use of the protocols and facilitate data retrieval, a comprehensive Excel® spreadsheet template was developed to include all the protocols (see Suppl. material 2). The resulting spreadsheet was checked by the authors or owners of each protocol to ensure that it accurately depicted the original protocol whilst matching the common-practice methodology.

Using the protocols selected in the spreadsheet template, 89 assessors independently assessed between three to 11 species (mean = 3.9) of the taxonomic group in their area of expertise (i.e. terrestrial plants, aquatic plants, terrestrial vertebrates, terrestrial insects, other terrestrial invertebrates, freshwater invertebrates, freshwater fish, marine species and pathogens) (Suppl. material 1: Table S1). All assessors were researchers with expertise in biological invasions (PhD or PhD candidate) selected among the participants of the Alien Challenge COST Action by the coordinators of each taxonomic group. The experience of the assessors with NNS impact assessments varied. Most assessors had occasionally participated in NNS risk assessments exercises (59.3%), while 19.7% had never participated and 17.5% had often participated. All NNS were assessed by a minimum of five assessors (maximum eight) (Suppl. material 1: Table S1), yielding a total of 2614 assessments. Before conducting the assessments, the assessors were required to read the impact assessment guidelines provided per protocol and ask

questions directly to the protocol developers if needed. When scoring impacts, assessors were instructed to consider Europe as the risk assessment area and the likely worst-case scenario for each NNS. Based on the precautionary principle, protocols recommend scoring the potential impact of NNS based on the available information either from studies for the area of assessment, or from areas with the same invaded habitat in a similar climate. The assessors were instructed to base their assessments on all available literature, information sources and their own expertise, indicating in the assessment the source of the information. The selection of the literature used for the assessment was left at the discretion of the assessor.

Before retrieving the data, each assessment was checked for completeness. Once all NNS assessments were completed, the final scores for each assessment were extracted. To harmonize scores across protocols, all ordinal scores (i.e. protocols with three or five levels as final scoring scale; Table 1) were transformed into numeric values, with the lowest impact as 1 and the maximum as 3 or 5, respectively. Then, all scores were standardized from 0 to 1 using the following equation $(S - Smin)/(Smax - Smin)$, where S represent the score per NNS in each assessment, and *Smax* and *Smin*, the maximum and minimum possible scores provided by the protocol (Turbé et al. 2017).

Consistency in non-native species scoring across assessors

For each NNS and protocol (471 combinations), the mean and the coefficient of variation (CV) of the final score were calculated. The mean was used as the overall score across experts per NNS and protocol, whereas CV was used as an estimate of the consistency of scores across experts, adjusting for the mean value. First, differences in CV among all protocols were tested using a linear mixed model with protocol name as a fixed effect and species nested within taxonomic groups as random effects (i.e. random intercept model). Second, we used multimodel inference (Burnham and Anderson 2002) of linear mixed models to analyze the relationship between the CV and species characteristics (taxonomic group and available knowledge), protocol characteristics (impact categories, spread question included, final scoring scale, whether final scoring was based on maximum score, number of questions and expertise on the species required) and assessor expertise on the species (mean and coefficient of variance). In this set of models, we used the same random effects structure as in the first model but did not include protocol name as a covariate. Model residuals were checked for normality and homoscedasticity and identified the square root as the best transformation for CV. Multi-model inference, based on the all-subsets selection of predictors, was performed using the corrected Akaike's Information Criterion (AIC_c) keeping the same random effects in all model combinations. For each combination of predictors, Akaike weights (w_i) were calculated. Considering the best models given the selected predictors ($\Delta AIC_c < 6$) (Richards 2008), the relative importance $w_{+(j)}$ of each predictor j was estimated as the sum of the AIC_c weights across all models in which the selected predictor appeared. Predictors with higher $w_{+(j)}$ (i.e. closer to 1) have a higher weight of evidence to explain

the response variable with the given data. Finally, the average of regression coefficients weighted by w_i within the subset of best models was calculated.

Differences in the mean CV among levels for the categorical variables in the best candidate model (i.e. with the smallest AIC_c) were tested for significance using a Tukey *post hoc* test. Prior to modelling, continuous predictors for the models above were checked for multicollinearity using Pearson correlations. All variables were selected for further analyses considering the low correlation values found ($r < 0.5$; Suppl. material 1: Table S2) (Dormann et al. 2013). Continuous variables were centered (deviate from the mean) and scaled (divided by standard deviation) to facilitate interpretation of model coefficients and model convergence (Schielzeth 2010). Finally, in all models explained above we accounted for the variability in the number of assessments per NNS (5 to 8; Suppl. material 1: Table S1) (i.e. sample size effect), including the number of assessments as a covariate (i.e. fixed effect).

Differences in impact assessment scoring across protocols

Similarities in the scoring of NNS across the different protocols were compared using hierarchical cluster analyses. Cluster analyses of the mean scores per NNS and protocol (calculations described above) were performed using Spearman's correlation coefficient as a similarity measure and the complete linkage method (i.e. maximum distance between clusters). Using this method, we first carried out a cluster analysis of all NNS across the six protocols common to all taxonomic groups (i.e. GABLIS, GB-NNRA, EICAT, Harmonia+, GISS and NGEIAAS). Then, separate analyses were also performed for four subsets of NNS with common protocols: 1) aquatic and terrestrial plants, 2) aquatic animals (combining freshwater invertebrates, freshwater fish, and marine invertebrates), 3) terrestrial invertebrates (terrestrial insects and other terrestrial invertebrates), and 4) terrestrial vertebrates (Suppl. material 1: Table S1). Pathogens were not included in this analysis due to the low number ($n = 3$) of species tested. Prior to these analyses in order to account for the variability in the number of assessments per NNS (five to eight; Suppl. material 1: Table S1) (i.e. sample size effect), we calculated the Pearson's correlation between the mean score per NSS and protocol and the number of assessments performed for all groups of species indicated above. When the correlation was significant for a group of species ($p < 0.05$) we used simple linear regression models to relate the mean score with the number of assessments per species and used the model's residuals in subsequent hierarchical analyses. We followed this approach only for plants and aquatic animals based on the significant correlation found (Plants r: -0.17, $p < 0.05$; Aquatic animals r: -0.17, $p < 0.05$). Results without this correction were similar, reinforcing the robustness of the results (Suppl. material 1: Fig. S2). All statistical analyses and figures were carried out in R v3.4.1 (R Core Team 2017) using packages lme4, lsmeans, MuMIn and sjPlot to implement and plot mixed models and gplots for the correlation heat maps and dendrograms.

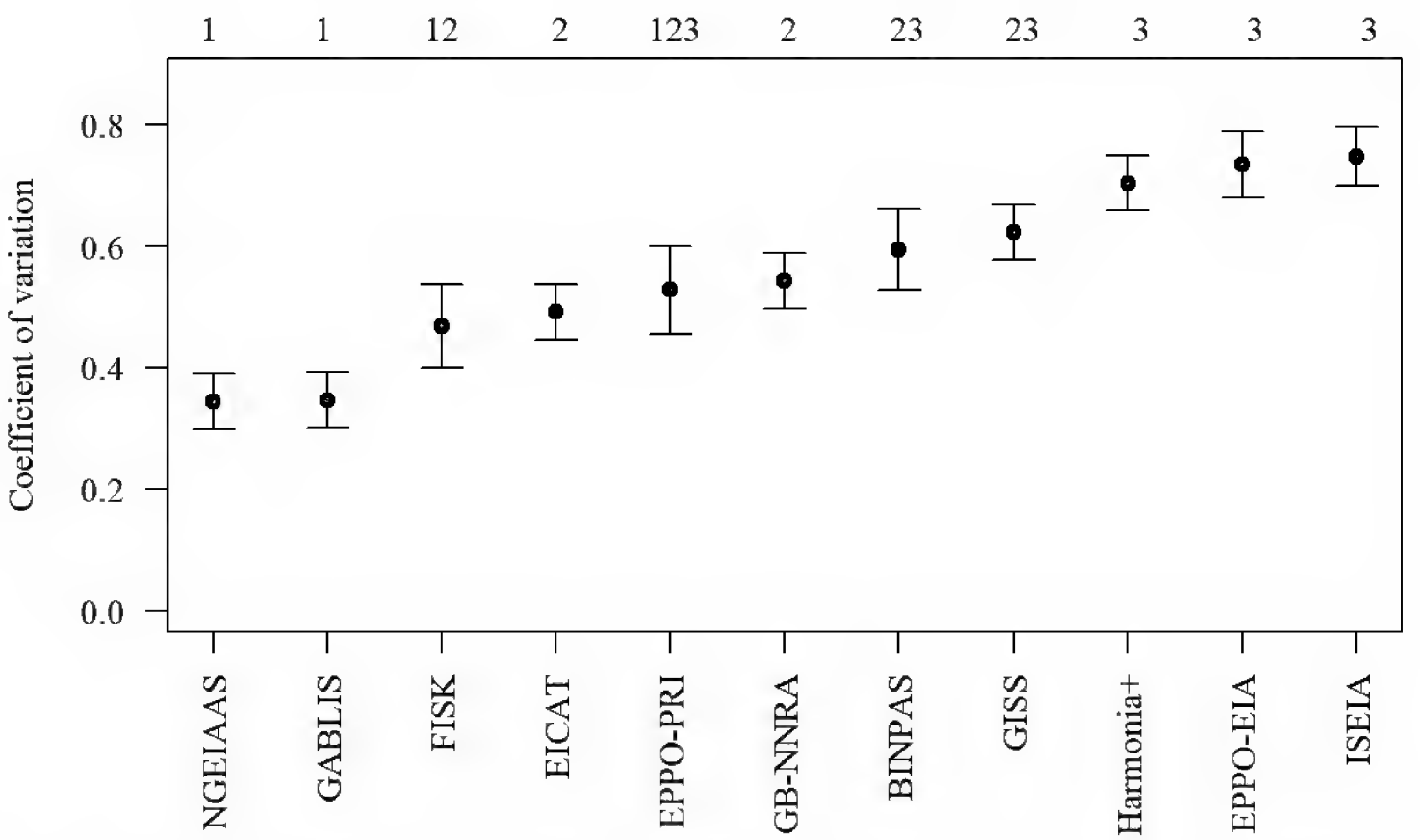

Figure 1. Coefficient of variation (CV) of species scoring across assessors per impact assessment protocol based on linear mixed models controlling for taxonomic group and species as nested random effects and number of assessments per species as fixed effects. Protocols with the same letters above the graph are not significantly different ($p < 0.05$; Tukey test). Dots indicate the least squares means per protocol. Lines indicate the confidence interval (95%) around the means.

Results

Consistency across assessors

The mean coefficient of variation (CV) of assessor scores per NNS and protocol was 40% (± 37% SD), with 10% ($n = 470$) showing complete agreement (CV = 0) among assessors but with maximum variability being 223% (four species in ISEIA: *Aedes albopictus*, *Arion vulgaris*, *Australoheros facetus* and *Fascioloides magna*; two species in EPPO EIA: *Diabrotica virgifera* and *Tuta absoluta*). CV was remarkably different among protocols (Fig. 1). ISEIA, EPPO-EIA and Harmonia+ protocols had the highest CV, whereas NGEIAAS and GABLIS protocols showed the lowest values. CV across assessors was better explained by protocol characteristics than by NNS characteristics (Table 2). Scoring scale, expertise required and the use of maximum impact score were the variables with the highest weight of evidence.

According to Tukey *post hoc* tests in the best candidate model, protocols using three score levels had significantly lower CV than the protocols using scales with five levels (difference = 0.25, $p < 0.001$) or more than five levels (difference = 0.29, $p < 0.001$). However, protocols with five score levels were similar to pro-

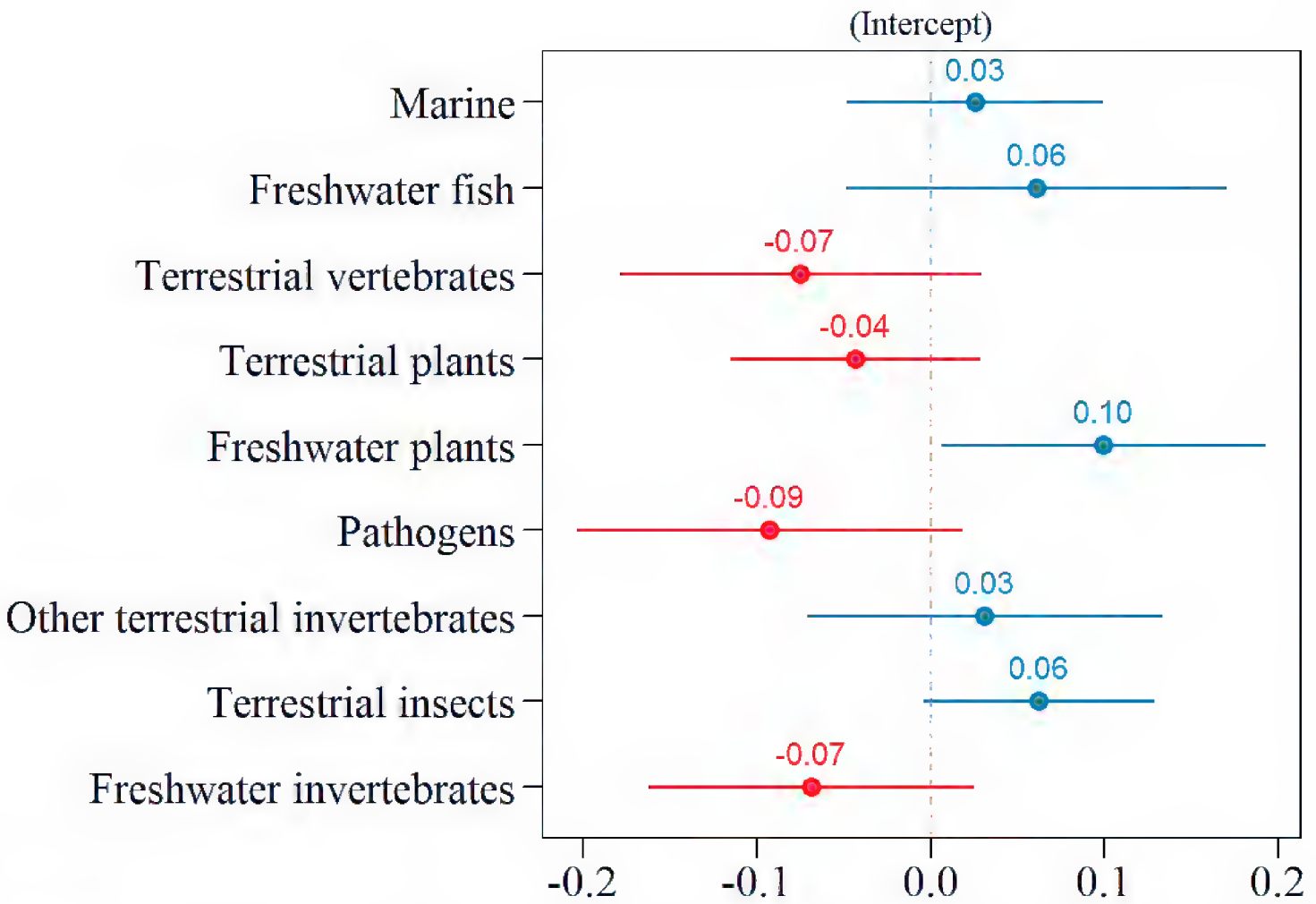

Figure 2. Mean regression coefficient and confidence interval (95%) of taxonomic groups (random effects) in the best linear mixed model explaining the coefficient of variation of scores of 57 invasive non-native species for 11 different protocols including all significant species, assessor and protocol characteristics (see Table 2) .

tocols with more than five levels (p = 0.27). CV across assessors was significantly lower for protocols that required higher expertise than those for which low expertise was required (Table 2). The expertise required per protocol was highly correlated to the overall number of fields in the protocol (i.e. questions, comments, uncertainty and results; Pearson's r = 0.9) but less with the number of questions actually contributing to the final score calculation (r = 0.5; Suppl. material 1: Table S2). Protocols using the maximum impact score yielded lower CV values. In terms of protocol content, CV was higher when protocols included a NNS spread module but there was no difference depending on the impact types considered (Table 2). The number of questions contributing to final score and impact categories considered did not show significant relations to CV (Table 2). Among NNS and assessor characteristics, only the mean of assessor expertise on each NNS showed a significant negative relationship with CV values (Table 2). Finally, there were some differences in CV among taxonomic groups (Fig. 2). Although not significant, terrestrial vertebrates, terrestrial plants, pathogens and freshwater invertebrates tended to show lower CVs whereas higher values were found for terrestrial insects, other terrestrial invertebrates and freshwater plants. Only terrestrial insects and freshwater plants showed a significantly higher CV than the average across all taxa (Fig. 2).

Table 2. Average coefficient and Akaike weights for each species, assessor and protocol variable within the best linear mixed models ($AIC_c < 6$) explaining the coefficient of variation of the scores of 57 non-native species in 11 impact assessment protocols. Taxonomic groups and species identification were included as nested random effect. Predictors with weight closer to one have a higher relative importance to explain the response variable. Variables with weight equals zero were not included in the best subset of models to calculate average coefficients.

Variable	Coefficient	Adjusted SE	*z*	*P*	Weight
Intercept	0.36	0.06	5.76	<0.001	
Number of assessments					0
Species					
Web of Science records (available knowledge)	-0.06	0.05	1.18	0.24	0.06
Assessor					
Mean assessor expertise	-0.04	0.02	2.21	0.03	0.14
CV assessor expertise					0
Protocol					
Scoring scale	See results section				1
Expertise required	-0.14	0.02	7.76	<0.001	1
Using maximum impact score (yes-no)	-0.12	0.02	4.93	<0.001	1
Spread (yes-no)	0.12	0.05	3.57	<0.001	0.95
Impact type					0
Number of questions					0

Consistency across protocols

The pair-wise correlations in NNS scores among the six protocols common to all taxa were highly diverse (min–max = 0.16–0.77; mean = 0.55), indicating low consistency in species scores among some protocols (Fig. 3). With respect to taxonomic groups, aquatic animals had the highest mean correlation among protocols, terrestrial invertebrates and plants showed an equally low mean correlation, and terrestrial vertebrates had the lowest correlation levels (Fig. 4). These correlations remained similar when considering only the protocols common to all three taxonomic groups (Suppl. material 1: Fig. S1) and without sample size correction (Suppl. material 1: Fig. S2). Cluster analysis identified two main groups (Fig. 3, Suppl. material 1: Fig. S1): protocols that include only environmental impacts (NGEIAAS, GABLIS, and EICAT) and protocols that include both environmental and socio-economic impacts (GB-NNRA, GISS and Harmonia+). The scorings of Harmonia+ were clearly distinct from the other protocols (indicated by lower correlation values), particularly for plants and terrestrial invertebrates (Figs 3, 4). Similarly, FISK and GABLIS showed relatively low correlation values with the other protocols for aquatic animals and terrestrial vertebrates, respectively (Fig. 4).

Discussion

The comparison of impact assessment protocols for NNS shows that scoring variability across assessors can be substantial, depending on the taxonomic group con-

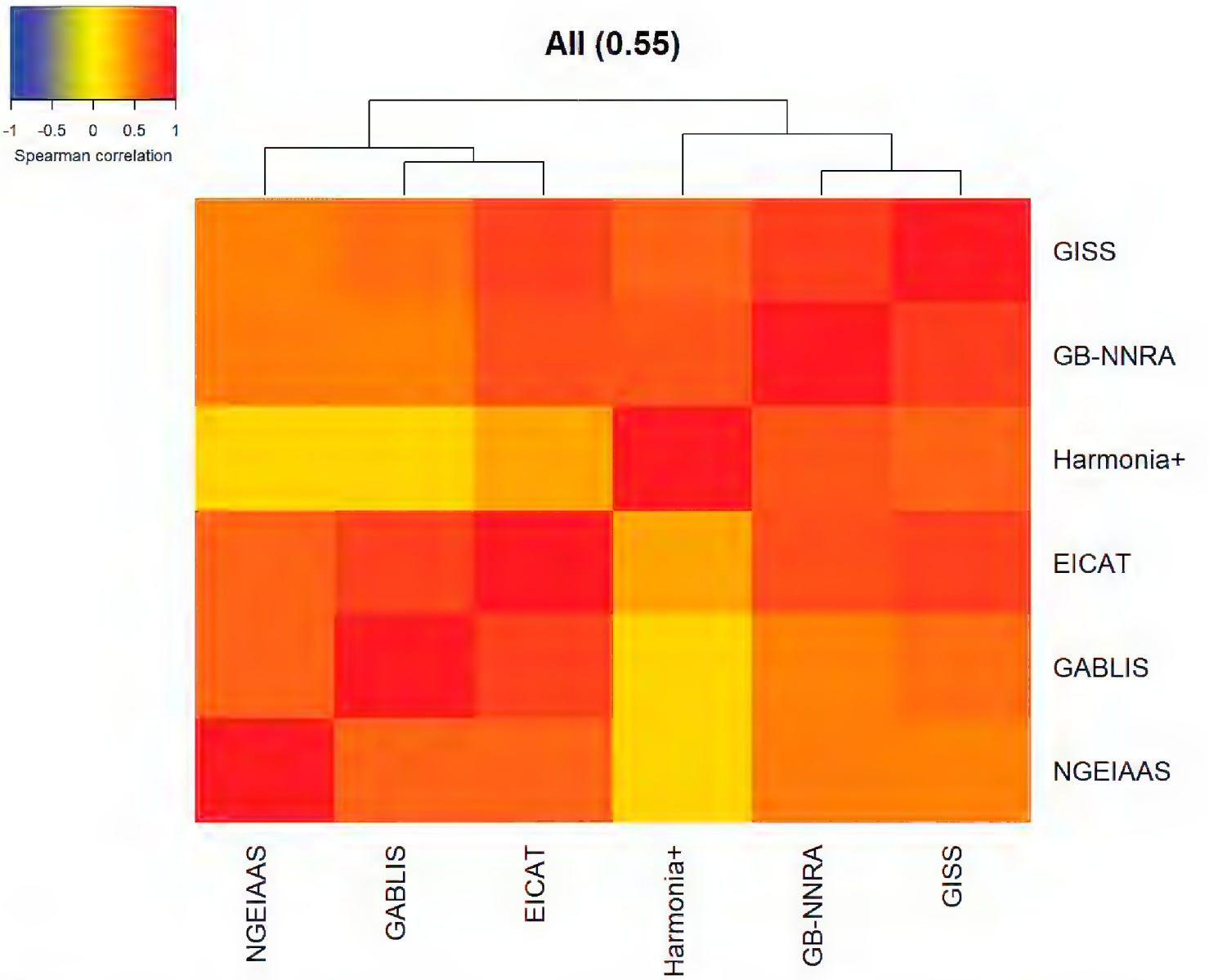

Figure 3. Spearman correlation matrix and hierarchical cluster of species scorings for the protocols common for all species. The color scale indicates the correlation between the species scorings obtained for each protocol pair. In brackets, the mean of all pair-wise correlations.

sidered and the scoring system. However, there is potential to reduce this variability by considering the expertise of the assessors and optimizing structural characteristics of the protocol. Furthermore, the ranking of NNS based on the protocol scoring can differ depending on the approach implemented, mainly based on the impact category type considered (i.e. whether socio-economic impacts are included). Thus, the selection of the scoring approach can have important consequences on the final ranking of NNS produced.

Consistency across assessors and across taxonomic groups

Scoring consistency across assessors and for some taxonomic groups was surprisingly low. It is not clear why these large discrepancies occurred even when the assessors were experts in invasion biology within their taxonomic domain. Many factors can influence the interpretations of context dependence found in the scientific literature, which can lead to subjective and inconsistent answers even amongst expert assessors (Gilovich

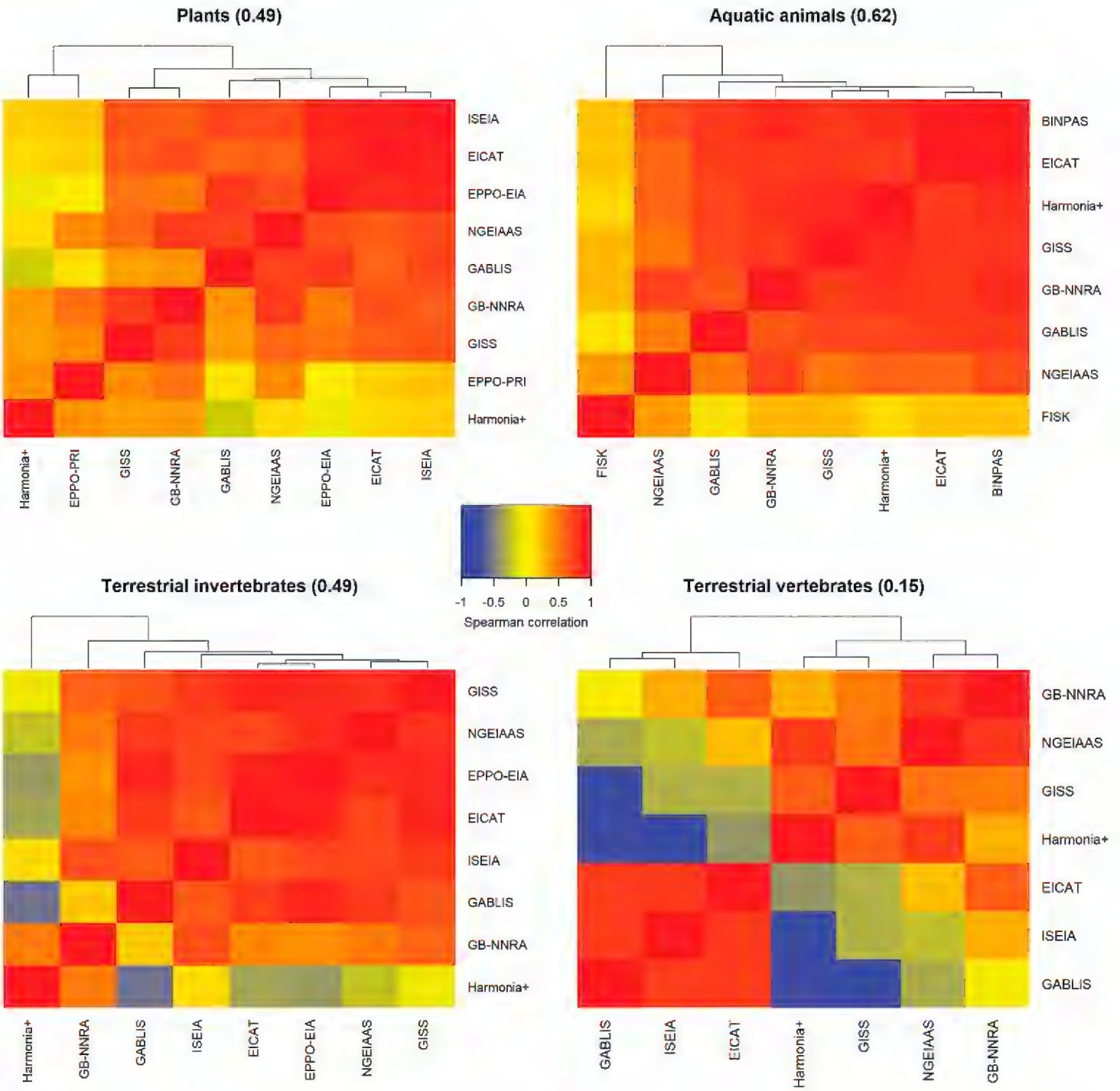

Figure 4. Spearman correlation matrix and hierarchical cluster of the species scorings for the protocols common per species group. The color scale indicates the correlation between the species scorings obtained for each protocol pair. In brackets, the mean of all pairwise correlations per group.

et al. 2002). Heuristics and bias, including intuitive strategies to process information, can lead to variability in expert responses (O'Hagan et al. 2006). For example, experts might score the impact according to the studies with which they feel most familiar (e.g. conducted by colleagues in their region). Similarly, if there is a lack of information on the impacts for a NNS, then the judgement might be biased towards a NNS of the same taxonomic lineage. Alternatively, inconsistencies might be due to inherent uncertainty. For instance, a greater inconsistency for most groups of aquatic taxa may reflect a higher difficulty in determining impacts than for taxa in other environments (Molnar et al. 2008). Finally, these biases could be balanced by anchoring effects where most assessors might assign intermediate levels of impact when there is insufficient information to fulfil the protocol requests.

Part of the variability in consistency was explained by protocol characteristics and the approaches implemented. Protocols with three score levels were more likely to show consistency among assessors than those with five or more levels. However, a three-category scoring system might not be sufficient to discriminate between NNS impacts or magnitude of impacts and rank NNS for prioritisation, because too many species will have the same score. Protocols that select the highest impact among different categories provided higher consistency. By definition, this approach will homogenise the scores towards higher values discarding inconsistencies from less important impacts in a way that results will be more conservative.

Protocols containing questions that required greater expertise on the species yielded higher scoring consistency than simpler protocols. Protocols requiring greater expertise demanded very detailed information about the species (e.g. expected population lifetime in NGEIAAS) that, when available, is very likely to be available only in few studies. Owing to the restricted number of sources of information, the variability in the final score might be low. Complex protocols might be less user-friendly and more time-consuming, but this in itself could increase focus and decrease subjectivity. Exceptions exist, e.g. the -ISK screening (Copp 2013), whereby the protocol is easy to use but the 49 questions require more time to answer than simpler tools such as ISEIA, which has only 12 questions. However, the questions from simple tools such as ISEIA focus mainly on impacts, whereas the -ISK screening tools include a much broader range of questions, such as invasion history, species traits and susceptibility to management measures. The balance between ease of use and time spent is critical as some protocols are meant to be used for the rapid screening of a NNS, whereas others provide more in-depth assessments. For example, NGEIAAS was designed for professional experts who carry out very detailed risk assessments on behalf of government authorities (Gederaas et al. 2012, Sandvik et al. 2013). This issue highlights that although we only selected impact and spread related sections, the present study compares tools intended for different phases of the risk analysis process, i.e. risk identification (e.g. ISEIA, -ISK screening tools), risk assessment (e.g. GB-NNRA, Harmonia+) and impact assessment (e.g. GISS, EICAT). Further studies could look into a detailed comparison across all phases of the risk analysis process in order to highlight those sections that might require improvement.

Regarding assessor and NNS characteristics, the only factor that significantly increased consistency among assessors was their level of expertise with the assessed species. Assessors that had previous experience with the NNS assessed may have had similar high levels of knowledge on that NNS, and this may have led to similar scores. Nevertheless, this situation is infrequent as NNS assessments are more commonly undertaken by persons familiar with the taxonomic group but not necessarily with the NNS being assessed (e.g. NNS not yet present or still rare in the study area). Unexpectedly, consistency was not related to the availability of information about the species (i.e. higher number of WoS records). The simplest explanation is that the number of studies available does not necessarily indicate more studies relevant for impact assessments as the literature on these species could be linked to other research fields in

invasion biology not directly associated with their environmental or socioeconomic impacts. It is also relevant to note that different assessors might have had access to different information sources, particularly non-English literature and reports. This might have affected consistency results but we followed standard practices for NNS risk assessments. Further studies could look at these differences providing a base information for the species to be assessed.

The high inconsistency found among assessor's scores raises high concerns and suggests that assessments conducted by single assessors should be interpreted with caution (Pheloung et al. 1999, Cousens 2008). Expert working group scoring, the use of consensus techniques and reviewing processes can inform the responses of single assessors and therefore reduce uncertainty (Sutherland and Burgman 2015, Vanderhoeven et al. 2017). For NNS lacking information or contrasting data, structured elicitation techniques, such as the Delphi approach, which is based on a feedback and revision process (Mukherjee et al. 2015), can identify and reduce potential sources of bias among experts (Morgan 2014, Sutherland and Burgman 2015). In practice, risk assessments for NNS, in particular those carried out in the plant health sector, are usually done either by groups of experts, as in EPPO pest risk assessment, or using an independent peer reviewer and an editorial-board type vetting procedure, such as in Great Britain (Baker et al. 2008, Booy et al. 2012). The consensus approach is used for plants and plant pests because those assessments are likely to be used in international trade agreements in order to demonstrate robustness (Schrader et al. 2010). However, national or regional impact risk assessments of NNS for blacklists or prioritization purposes are often based on the judgement of a few or single experts. Thus, efforts should be made to involve a panel of experts in the species or the system following elicitation techniques.

Differences across protocols

Variations among protocols in species scoring are mainly due to the inclusion, or not, of socio-economic impacts. Although socio-economic and environmental impacts are generally correlated (Kumschick et al. 2015a, Vilà et al. 2010), it is almost impossible to predict the magnitude of one impact from the other (Bacher et al. 2018). For example, many NNS, such as agricultural pests and organisms affecting human health, exclusively cause socio-economic impacts (Kenis and Branco 2010, Kumschick et al. 2015b) and, thus, using protocols that include such impacts will affect the impact ranking of NNS under consideration. Furthermore, the perception of socio-economic impacts is likely to vary across stakeholders. Thus, depending on the target audience and objectives of the assessment, different protocols may be used, focusing either on environmental or socio-economic impacts or both together. The majority of the protocols exclusively considered environmental impacts, and there was greater correlation in scores among these protocols. However, the difference between scores was dependent on the taxonomic group under consideration. Ranking of species completely shifted (negative correlation of scores across protocols) when different impact categories were

considered for terrestrial vertebrates and plants, but the difference was lower for aquatic animals. This pattern might be due to differences in the relevance of impacts across taxa, with terrestrial vertebrates showing highly contrasting impact types for single species (e.g. high economic impact but low environmental impact) (Vilà et al. 2010). However, differences in scores among taxonomic groups might again also simply reflect differences in the knowledge of their impacts. Impacts of terrestrial vertebrates or plants might be better known than those of aquatic organisms. Testing this hypothesis requires comparing uncertainty scores provided by experts across impact types and taxonomic groups, which could be done with the current dataset in further studies.

Among all protocols, Harmonia+, FISK and GABLIS led to very different scores in comparison to the other protocols. This difference was partly related to the different impact categories considered but also to the inclusion of questions beyond impact (e.g. management in GABLIS and FISK). Finally, the GB-NNRA protocol showed a variable relation with other protocols across taxa: low correlation with protocols only considering environmental impacts for plants and terrestrial invertebrates but high for vertebrates. The final score in the GB-NNRA was not automatically calculated as in the other protocols. Instead, assessors were asked to provide overall summary scores and confidence rankings for the NNS based on the answers provided in previous sections, which include questions that consider both environmental and socio-economic impacts (Baker et al. 2008, Mumford et al. 2010). This approach could have led to the not consistent relation between the GB-NNRA protocol and the others. However, when used as part of the GB risk analysis process (Booy et al. 2012), it aids the NNS risk analysis panel to identify inconsistencies between the assessor's individual question responses and their overall scores and confidence levels (Mumford et al. 2010).

Recommendations for NNS impact assessments

Several key factors should be taken into account when selecting or designing a NNS risk assessment protocol, such as the aim, the scope, the consistency and the accuracy of the outcomes, and the resources available to perform the assessment (e.g. time or information). As a first step, the suitability of a NNS risk assessment protocol will depend on the scope and aim of the assessment. For instance, if a NNS is already present in the region of interest, assessments on likelihood of entry and establishment are less meaningful than just the assessment of impact. Protocols with different scopes may produce different results in terms of NNS rankings (Lazzaro et al. 2016). As we have shown, even just considering different types of impacts could result in large differences in rankings. Thus, it is crucial not to mix the results from assessment methods that consider different impacts or phases of the invasion process. Furthermore, our results show that even if the focus is only on impact and spread sections, the choice of the protocol is critical because the scoring consistency will depend on the characteristics of the protocol. Three main factors were responsible for these inconsistencies, the choice of the scoring scale, how the final score is summarized and the general expertise required to use the

protocol. We recommend using a 5-level scoring, maximum aggregation method and moderate expertise requirements as a good compromise to reduce inconsistency without losing discriminatory power or usability. In general, we also advise protocol developers to perform sensibility tests of consistency before final release or adoption (e.g. Pheloung et al. 1999). This is crucial because if a protocol yields inconsistent outcomes when used by different assessors, then it is likely that decisions taken based on the results could be variable and disproportionate to the actual impacts (Schrader et al. 2012).

Part of the inconsistency might also come from the way the protocol is used in practice (e.g. standardized forms, clear guidelines, selection of assessors, individual vs. group assessments). We propose three main ways to reduce this type of inconsistency. First, irrespectively of the protocol, selecting a group of assessors with high expertise will yield more consistent results. Second, inconsistencies due to linguistic uncertainties (e.g. definitions, formulations, rating) can be reduced by improving the guidelines and with adequate training of the assessors (Vilà et al. 2019). Third, other studies have suggested using expert elicitation methods to reduce inconsistencies (Morgan 2014, Sutherland and Burgman 2015), such as consensus building (Mukherjee et al. 2015) or quality control mechanisms (e.g. peer-review panels). Elicitation methods can reveal whether differences in scoring outcomes between and within protocols reflect true differences in opinion, lack of evidence, or subjective biases due to protocol interpretation (Vanderhoeven et al. 2017). In fact, scientific consensus and robust revisions are crucial for policy implementation (Turbé et al. 2017). Finally, there will always be inconsistencies due to knowledge gaps and subjectivity in the interpretation of the scientific results when there is high context dependency. This might not be a problem in providing a sound evidence-base for decisions on NNS as long as protocols are used transparently and uncertainties are explicitly dealt with through appropriate methods (Vanderhoeven et al. 2017).

Acknowledgements

This article is based upon work from the COST Action TD1209: Alien Challenge. COST (European Cooperation in Science and Technology) is a pan-European inter-governmental framework. The mission of COST is to enable scientific and technological developments leading to new concepts and products and thereby contribute to strengthening Europe's research and innovation capacities. PGM was supported by the CABI Development Fund (with contributions from ACIAR (Australia) and Dfid (UK) and by Darwin plus, DPLUS074 'Improving biosecurity in the SAUKOTs through Pest Risk Assessments'. MV by Belmont Forum-Biodiversa project InvasiBES (PCI2018-092939). CP by Sciex-NMSch 12.108. JMJ and WCS by BiodivERsA (FFII project; DFG grant JE 288/7-1). JMJ by DFG project JE 288/9-1,9-2. CR and MB by Fundação para a Ciência e a Tecnologia grants SFRH/BPD/91357/2012 and SFRH/BPD/86215/2012, respectively. PS by MESTD of Serbia, grant #173025. JP by RVO 67985939 and 17-19025S. JCC was supported by a starting grant in the framework of the 2014 FCT Investigator Programme (IF/01606/2014/CP1230/CT0001).

References

Bacher S, Blackburn TM, Essl F, Genovesi P, Heikkilä J, Jeschke JM, Jones G, Keller R, Kenis M, Kueffer C, Martinou AF, Nentwig W, Pergl J, Pyšek P, Rabitsch W, Richardson DM, Roy HE, Saul W-C, Scalera R, Vilà M, Wilson JRU, Kumschick S (2018) Socio-economic impact classification of alien taxa (SEICAT). Methods in Ecology and Evolution 9: 159–168. https://doi.org/10.1111/2041-210X.12844

Baker R, Black R, Copp GH, Haysom K, Hulme PE, Thomas M, Ellis M (2008) The UK risk assessment scheme for all non-native species. In: Rabitsch W, Essl F, Klingenstein F (Eds) Biological invasions: from ecology to conservation. Institute of Ecology of the TU Berlin, Berlin, 46–57.

Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JRU, Richardson DM (2011) A proposed unified framework for biological invasions. Trends in Ecology & Evolution 26: 333–339. https://doi.org/10.1016/j.tree.2011.03.023

Booy O, Copp GH, Mazaubert É (2012) Réseaux d'experts et prise de décisions: l'exemple du Royaume-Uni. Sciences Eaux & Territoires Numéro 6: 74–77.

Branquart E (2009) Guidelines for environmental impact assessment and list classification of non-native organisms in Belgium. Version 2.6.

Brunel S, Branquart E, Fried G, Van Valkenburg J, Brundu G, Starfinger U, Buholzer S, Uludag A, Joseffson M, Baker R (2010) The EPPO prioritization process for invasive alien plants. EPPO Bulletin 40: 407–422. https://doi.org/10.1111/j.1365-2338.2010.02423.x

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approachSecond edition. Springer-Verlag, New York, 528 pp.

Copp GH (2013) The Fish Invasiveness Screening Kit (FISK) for non-native freshwater fishes: A summary of current applications. Risk Analysis 33: 1394–1396. https://doi.org/10.1111/risa.12095

Copp GH, Vilizzi L, Mumford J, Fenwick GV, Godard MJ, Gozlan RE (2009) Calibration of FISK, an Invasiveness Screening Tool for Nonnative Freshwater Fishes. Risk Analysis 29: 457–467. https://doi.org/10.1111/j.1539-6924.2008.01159.x

Copp GH, Vilizzi L, Tidbury H, Stebbing PD, Trakan AS, Miossec L, Goulletquer P (2016) Development of a generic decision-support tool for identifying potentially invasive aquatic taxa: AS-ISK. Management of Biological Invasions 7: 343–350. https://doi.org/10.3391/mbi.2016.7.4.04

Cousens R (2008) Risk assessment of potential biofuel species: an application for trait-based models for predicting weediness? Weed Science 56: 873–882. https://doi.org/10.1614/WS-08-047.1

D'hondt B, Vanderhoeven S, Roelandt S, Mayer F, Versteirt V, Adriaens T, Ducheyne E, Martin GS, Grégoire J-C, Stiers I, Quoilin S, Cigar J, Heughebaert A, Branquart E (2015) Harmonia+ and Pandora+: risk screening tools for potentially invasive plants, animals and their pathogens. Biological Invasions 17: 1869–1883. https://doi.org/10.1007/s10530-015-0843-1

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: a review of methods to deal

with it and a simulation study evaluating their performance. Ecography 36: 027–046. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Essl F, Nehring S, Klingenstein F, Milasowszky N, Nowack C, Rabitsch W (2011) Review of risk assessment systems of IAS in Europe and introducing the German-Austrian Black List Information System (GABLIS). Journal for Nature Conservation 19: 339–350. https://doi.org/10.1016/j.jnc.2011.08.005

Gederaas L, Moen TL, Skjelseth S, Larsen LK (2012) Non-native species in Norway – with the Norwegian Black List 2012. The Norwegian Biodiversity Information Centre, Norway.

Gilovich T, Griffin D, Kahneman D (2002) Heuristics and biases: The psychology of intuitive judgment. Cambridge University Press, Cambridge.

Hawkins CL, Bacher S, Essl F, Hulme PE, Jeschke JM, Kühn I, Kumschick S, Nentwig W, Pergl J, Pyšek P, Rabitsch W, Richardson DM, Vilà M, Wilson JRU, Genovesi P, Blackburn TM (2015) Framework and guidelines for implementing the proposed IUCN Environmental Impact Classification for Alien Taxa (EICAT). Diversity and Distributions 21: 1360–1363. https://doi.org/10.1111/ddi.12379

Heikkilä J (2011) A review of risk prioritisation schemes of pathogens, pests and weeds: principles and practices. Agricultural and Food Science 20: 15–28. https://doi.org/10.2137/145960611795163088

Katsanevakis S, Wallentinus I, Zenetos A, Leppäkoski E, Çinar ME, Oztürk B, Grabowski M, Golani D, Cardoso AC (2014) Impacts of invasive alien marine species on ecosystem services and biodiversity: a pan-European review. Aquatic Invasions 9: 391–423. https://doi.org/10.3391/ai.2014.9.4.01

Kenis M, Bacher S, Baker RHA, Branquart E, Brunel S, Holt J, Hulme PE, MacLeod A, Pergl J, Petter F, Pyšek P, Schrader G, Sissons A, Starfinger U, Schaffner U (2012) New protocols to assess the environmental impact of pests in the EPPO decision-support scheme for pest risk analysis. EPPO Bulletin 42: 21–27. https://doi.org/10.1111/j.1365-2338.2012.02527.x

Kenis M, Branco M (2010) Impact of alien terrestrial arthropods in Europe. Chapter 5. BIORISK? Biodiversity and Ecosystem Risk Assessment 4. https://doi.org/10.3897/biorisk.4.42

Křivánek M, Pyšek P (2006) Predicting invasions by woody species in a temperate zone: a test of three risk assessment schemes in the Czech Republic (Central Europe). Diversity and Distributions 12: 319–327. https://doi.org/10.1111/j.1366-9516.2006.00249.x

Kumschick S, Bacher S, Evans T, Marková Z, Pergl J, Pyšek P, Vaes-Petignat S, van der Veer G, Vilà M, Nentwig W (2015a) Comparing impacts of alien plants and animals in Europe using a standard scoring system. Journal of Applied Ecology 52: 552–561. https://doi.org/10.1111/1365-2664.12427

Kumschick S, Gaertner M, Vilà M, Essl F, Jeschke JM, Pyšek P, Ricciardi A, Bacher S, Blackburn T, Dick J, Evans T, Hulme PE, Kühn I, Mrugała A, Pergl J, Rabitsch W, Richardson D, Sendek A, Winter M (2015b) Ecological Impacts of Alien Species: Quantification, Scope, Caveats, and Recommendations. BioScience 65. https://doi.org/10.1093/biosci/biu193

Lazzaro L, Foggi B, Ferretti G, Brundu G (2016) Priority invasive alien plants in the Tuscan Archipelago (Italy): comparing the EPPO prioritization scheme with the Australian WRA. Biological Invasions 18: 1317–1333. https://doi.org/10.1007/s10530-016-1069-6

Leung B, Roura-Pascual N, Bacher S, Heikkilä J, Brotons L, Burgman MA, Dehnen-Schmutz K, Essl F, Hulme PE, Richardson DM, Sol D, Vilà M (2012) TEASIng apart alien species risk assessments: a framework for best practices. Ecology Letters 15: 1475–1493. https://doi.org/10.1111/ele.12003

Matthews J, van der Velde G, Collas FPL, de Hoop L, Koopman KR, Hendriks AJ, Leuven RSEW (2017) Inconsistencies in the risk classification of alien species and implications for risk assessment in the European Union. Ecosphere 8: 1–17. https://doi.org/10.1002/ecs2.1832

McGeoch MA, Genovesi P, Bellingham PJ, Costello MJ, McGrannachan C, Sheppard A (2016) Prioritizing species, pathways, and sites to achieve conservation targets for biological invasion. Biological Invasions 18: 299–314. https://doi.org/10.1007/s10530-015-1013-1

Molnar JL, Gamboa RL, Revenga C, Spalding MD (2008) Assessing the global threat of invasive species to marine biodiversity. Frontiers in Ecology and the Environment 6: 485–492. https://doi.org/10.1890/070064

Morgan MG (2014) Use (and abuse) of expert elicitation in support of decision making for public policy. Proceedings of the National Academy of Sciences 111: 7176–7184. https://doi.org/10.1073/pnas.1319946111

Mukherjee N, Hugé J, Sutherland WJ, McNeill J, Van Opstal M, Dahdouh-Guebas F, Koedam N (2015) The Delphi technique in ecology and biological conservation: applications and guidelines. Methods in Ecology and Evolution 6: 1097–1109. https://doi.org/10.1111/2041-210X.12387

Mumford JD, Booy O, Baker RHA, Rees M, Copp GH, Black K, Holt J, Leach AW, Hartley M (2010) Invasive non-native species risk assessment in Great Britain. Aspects of Applied Biology, 49–54.

Narščius A, Olenin S, Zaiko A, Minchin D (2012) Biological invasion impact assessment system: From idea to implementation. Ecological Informatics 7: 46–51. https://doi.org/10.1016/j.ecoinf.2011.11.003

Nentwig W, Bacher S, Pyšek P, Vilà M, Kumschick S (2016) The generic impact scoring system (GISS): a standardized tool to quantify the impacts of alien species. Environmental Monitoring and Assessment 188: 315. https://doi.org/10.1007/s10661-016-5321-4

Nentwig W, Kühnel E, Bacher S (2010) A Generic Impact-Scoring System Applied to Alien Mammals in Europe. Conservation Biology 24: 302–311. https://doi.org/10.1111/j.1523-1739.2009.01289.x

O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ, Oakley JE, Rakow T (2006) Uncertain Judgements: Eliciting Experts' Probabilities. John Wiley & Sons, Chichester, 340 pp. https://doi.org/10.1002/0470033312

Olenin S, Minchin D, Daunys D (2007) Assessment of biopollution in aquatic ecosystems. Marine Pollution Bulletin 55: 379–394. https://doi.org/10.1016/j.marpolbul.2007.01.010

Panov VE, Alexandrov B, Arbačiauskas K, Binimelis R, Copp GH, Grabowski M, Lucy F, Leuven RS, Nehring S, Paunović M (2009) Assessing the risks of aquatic species invasions via European inland waterways: from concepts to environmental indicators. Integrated environmental assessment and management 5: 110–126. https://doi.org/10.1897/IEAM_2008-034.1

Pheloung PC, Williams PA, Halloy SR (1999) A weed risk assessment model for use as a biosecurity tool evaluating plant introductions. Journal of environmental management 57: 239–251. https://doi.org/10.1006/jema.1999.0297

R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Richards SA (2008) Dealing with overdispersed count data in applied ecology. Journal of Applied Ecology 45: 218–227. https://doi.org/10.1111/j.1365-2664.2007.01377.x

Roy HE, Rabitsch W, Scalera R, Stewart A, Gallardo B, Genovesi P, Essl F, Adriaens T, Bacher S, Booy O, Branquart E, Brunel S, Copp GH, Dean H, D'hondt B, Josefsson M, Kenis M, Kettunen M, Linnamagi M, Lucy F, Martinou A, Moore N, Nentwig W, Nieto A, Pergl J, Peyton J, Roques A, Schindler S, Schönrogge K, Solarz W, Stebbing PD, Trichkova T, Vanderhoeven S, van Valkenburg J, Zenetos A (2018) Developing a framework of minimum standards for the risk assessment of alien species. Journal of Applied Ecology 55: 526–538. https://doi.org/10.1111/1365-2664.13025

Sandvik H, Sæther B-E, Holmern T, Tufto J, Engen S, Roy HE (2013) Generic ecological impact assessments of alien species in Norway: a semi-quantitative set of criteria. Biodiversity and Conservation 22: 37–62. https://doi.org/10.1007/s10531-012-0394-z

Schielzeth H (2010) Simple means to improve the interpretability of regression coefficients. Methods in Ecology and Evolution 1: 103–113. https://doi.org/10.1111/j.2041-210X.2010.00012.x

Schrader G, MacLeod A, Mittinty M, Brunel S, Kaminski K, Kehlenbeck H, Petter F, Baker R (2010) Enhancements of pest risk analysis techniques. EPPO Bulletin 40: 107–120. https://doi.org/10.1111/j.1365-2338.2009.02360.x

Schrader G, MacLeod A, Petter F, Baker RHA, Brunel S, Holt J, Leach AW, Mumford JD (2012) Consistency in pest risk analysis – how can it be achieved and what are the benefits? EPPO Bulletin 42: 3–12. https://doi.org/10.1111/j.1365-2338.2012.02547.x

Snyder E, Mandrak NE, Niblock H, Cudmore B (2013) Developing a Screening Level Risk Assessment Prioritization Protocol for Aquatic Non-Indigenous Species in Canada: Review of Existing Protocols. Fisheries and Oceans Canada, Canadian Science Advisory Secretariat: Research Document 97: 1–82.

Sutherland WJ, Burgman M (2015) Policy advice: Use experts wisely. Nature 526: 317–318. https://doi.org/10.1038/526317a

Tricarico E, Vilizzi L, Gherardi F, Copp GH (2010) Calibration of FI-ISK, an Invasiveness Screening Tool for Nonnative Freshwater Invertebrates. Risk Analysis 30: 285–292. https://doi.org/10.1111/j.1539-6924.2009.01255.x

Turbé A, Strubbe D, Mori E, Carrete M, Chiron F, Clergeau P, González-Moreno P, Le Louarn M, Luna A, Menchetti M, Nentwig W, Pârâu LG, Postigo J-L, Rabitsch W, Senar JC, Tollington S, Vanderhoeven S, Weiserbs A, Shwartz A (2017) Assessing the assessments: evaluation of four impact assessment protocols for invasive alien species. Diversity and Distributions 23(3): 297–307. https://doi.org/10.1111/ddi.12528

Vanderhoeven S, Branquart E, Casaer J, D'hondt B, Hulme PE, Shwartz A, Strubbe D, Turbé A, Verreycken H, Adriaens T (2017) Beyond protocols: improving the reliability of expert-based risk analysis underpinning invasive species policies. Biological Invasions 19(9): 2507–2517. https://doi.org/10.1007/s10530-017-1434-0

Vilà M, Basnou C, Pyšek P, Josefsson M, Genovesi P, Gollasch S, Nentwig W, Olenin S, Roques A, Roy D, Hulme PE (2010) How well do we understand the impacts of alien species on ecosystem services? A pan-European, cross-taxa assessment. Frontiers in Ecology and the Environment 8: 135–144. https://doi.org/10.1890/080083

Vilà M, Espinar JL, Hejda M, Hulme PE, Jarošík V, Maron JL, Pergl J, Schaffner U, Sun Y, Pyšek P (2011) Ecological impacts of invasive alien plants: a meta-analysis of their effects on species, communities and ecosystems. Ecology letters 14: 702–708. https://doi.org/10.1111/j.1461-0248.2011.01628.x

Vilà M, Gallardo B, Preda C, García-Berthou E, Essl F, Kenis M, Roy HE, González-Moreno P (2019) A review of impact assessment protocols of non-native plants. Biological Invasions 21: 709–723. https://doi.org/10.1007/s10530-018-1872-3

Vilà M, Hulme PE (2017) Impact of Biological Invasions on Ecosystem Services. Springer International Publishing, Cham, 354 pp.

Zaiko A, Lehtiniemi M, Narščius A, Olenin S (2011) Assessment of bioinvasion impacts on a regional scale: a comparative approach. Biological Invasions 13: 1739–1765. https://doi.org/10.1007/s10530-010-9928-z

Supplementary material 1

Supplementary materials

Authors: Pablo González-Moreno et al.

Data type: statistical data

Explanation note: Figure S1: hierarchical cluster of the species scores for the six protocols common to all taxonomic groups. Figure S2: hierarchical cluster of the species scorings for plants and aquatic animals without correcting for sample size bias. Table S1: list of non-native species. Table S2: correlation analyses.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.44.31650.suppl1

Supplementary material 2

Supplementary materials

Authors: Pablo González-Moreno et al.

Data type: Spreadsheet template

Explanation note: Spreadsheet template to fill the 11 impact assessment protocols for non-native species considered in the study.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.44.31650.suppl2